**Article**

# Calibrating the Impact of Vision Impairment (IVI): Creation of a Sample-Independent Visual Function Measure for Patient-Centered Outcomes Research

Judith E. Goldstein[1], Eva Fenwick[2], Robert P. Finger[3], Vijaya Gothwal[4], Mary Lou Jackson[5], Ecosse Lamoureux[2], Gwyneth Rees[6], and Robert Massof[1]

[1] Wilmer Eye Institute, Johns Hopkins University, Baltimore, MD, USA
[2] Population Health, Singapore Eye Research Institute, Singapore, Singapore
[3] University of Bonn, Bonn, Germany
[4] Deshpande Centre for Sight Enhancement Brien Holden Eye Research Centre – Patient Reported Outcomes Unit, Hyderabad, India
[5] University of British Columbia Department of Ophthalmology and Visual Sciences, Vancouver, BC, Canada
[6] University of Melbourne, Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, Australia

**Purpose:** Provide item calibrations estimated for the Impact of Vision Impairment (IVI) questionnaire by pooling data from several studies of people with vision impairment (VI) representing a variety of countries and causes of VI.

**Methods:** Eight data sets from six principal investigators representing responses to IVI items from 2867 VI patients were pooled for analysis. Eligible patients were 18 years or older and from Australia, India, and the United States. Rasch analysis, using the Andrich Rating Scale Model (Winsteps version 3.65), was performed on preintervention IVI responses to estimate item and person measures, reliability coefficients, and response category thresholds. Differential item functioning (DIF) analysis and analysis of variance (ANOVA) were used to examine the effects different data sets and covariates on item estimates.

**Results:** Patient age range was 18 to 103 years (median 62 years); 55% were male. Visual acuity ranged from 20/20 to no light perception and primary diagnosis was macular degeneration in 29% of patients. Item measure estimates showed good separation reliability ($R^2 = 0.99$). DIF magnitude did not preclude use of all IVI-28 data. ANOVA showed VA ($P < 0.001$) and gender ($P < 0.002$) were predictors of visual ability.

**Conclusions:** Analysis from pooled data support the provision of calibrated IVI item measures for researchers and clinicians to use, thus better enabling direct comparisons of patients with VI.

**Translational Relevance:** Validity testing of the IVI show that we can combine disparate data sets of patient responses to calibrate item measures and response category thresholds, and provide to others for use in comparing patients across clinical trials and on an individual basis.

## Introduction

Patient-reported outcome measures (PROMs) can be used to estimate the effects of treatment from the patient's perspective. In the form of visual function questionnaires (VFQ), PROMs are included routinely in ophthalmic clinical trials. There are numerous VFQs from which to choose.[1–4] Many VFQs still used today were developed before the adoption of modern psychometric methods by clinical vision researchers. Consequently, legacy VFQs are not calibrated and published measurement scales are specific to the instrument and patient samples. With increasing interest in patient-centered outcomes, regulatory agencies, research funding institutions, and third-

party payers are recognizing the need for standardization of PROMs so they can be applied meaningfully to a patient population.[5–7] Standardization requires that the same measurement scale be used across instruments for a patient population so that outcomes can be compared between studies.[8,9]

Most VFQs are designed to measure visual ability, a latent trait that refers to the patient's ability to perform activities that depend on vision. With an increasing number of ophthalmic interventions that may not be restorative, but rather are intended to slow the rate of visual decline (e.g., anti-VEGF therapy in age-related macular degeneration [AMD]) or maximize functional vision (e.g., Implantable Miniature Telescope, Argus II, etc.), accurate and precise measures of patient-reported visual ability outcomes are necessary for patient-centered comparative effectiveness research.[10,11]

VFQs consist of a set of items (i.e., questions) that describe vision-dependent activities requiring varying levels of ability (e.g., operating the telephone, reading labels on medicine bottles) and a set of ordered rating categories that increase in response magnitude (e.g., not at all difficult, moderately difficult, a lot of difficulty). Generating item responses from a VFQ for a well-defined group of patients and applying Rasch analysis results in estimates of a single measure for each respondent. This "person measure" refers to the magnitude of the patient's latent trait that one wants to quantify (e.g., visual ability). Item measures are on the same scale as, but independent of, person measures and refer to the magnitude of the latent trait a person must have to endorse the item at a particular level (e.g., the minimum visual ability a person must have to perform the described activity with a criterion level of ease). The practice in past studies has been to estimate measures such as item, person, and rating category threshold (point where the probability of responding to one category equals the probability of responding to the adjacent category) simultaneously from the study data matrix using a joint maximum likelihood estimation procedure. Since these measures are independent, but subject to imprecision and bias with small samples, and the logit scale estimated from Rasch models is dimensionless and normalized to the intrinsic uncertainty in the estimate, the ideal is to estimate item and rating category threshold measures from a large sample that is representative of the patient population to which the measure will be applied and then future estimates of person measures would be made with the item measures and rating category thresholds anchored to the calibrated values.[12] In that case, like using a ruler, person measures can be estimated from respondent ratings one person at a time, with no loss of precision or accuracy.

The 28-item Impact of Visual Impairment (IVI) Questionnaire is one of the more commonly used VFQs in ophthalmic studies outside the United States. It is relatively short and easy to administer. The IVI was designed to assess participation in vision-dependent daily activities and has been used to assess the effectiveness of ophthalmic treatment and low vision rehabilitation.[13–18] As part of the original work, factor analysis suggested there are three separate subscales on the IVI VFQ, "Reading and Accessing Information" (nine items), "Mobility and Independence" (11 items), and "Emotional Well-being" (eight items).[18]

The aim of this report is to provide item calibrations estimated for the IVI by pooling data from several studies of people with vision impairment representing a variety of countries and causes of visual impairment. An analysis of pooled data will improve our understanding of the underlying properties of the IVI and provide item calibrations that can consistently be applied in future studies enabling direct comparisons across studies of patient-reported outcomes of clinical research on individuals with vision impairment.

## Methods

All studies that provided data for this analysis had institutional review board (IRB) approval and adhered to the tenets of the Declaration of Helsinki. Analysis of the de-identified data provided by study collaborators did not require further approval by Johns Hopkins IRB.

### Participants and Design

Raw data from eight patient samples from previously published studies were used in this analysis (Table 1). In all samples, eligible patients were age 18 years or older and identified as low vision patients. Study participants were from Australia, India, and the United States and most questionnaire administrations were performed in English. For many participants in India, the questionnaire was administered in Telugu or Hindi.[19] Data from each sample included baseline raw scores of patient responses to the IVI, a key for the item description and item response, visual acuity (VA) for each eye or for

**Table 1.** Patient Characteristics

| | Patient Characteristics of IVI 28 Data Sets | | | | |
|---|---|---|---|---|---|
| | Rees-1 IVI 28 (n = 153) | Rees-2 IVI 28 (n = 168) | Finger IVI 28 (n = 597) | Gothwal IVI 28 (n = 1141) | Jackson IVI 28 (n = 267) |
| Gender (%) | | | | | |
| Male | 39.87 | 32.74 | 46.70 | 75.48% | 44.53% |
| Female | 60.13 | 67.26 | 53.30 | 24.52% | 55.47% |
| VA[a] | | | | | |
| logMAR 0–0.3 | 0.00 | 4.27 | 0.00 | 20.58% | 39.62% |
| logMAR 0.31–0.5 | 16.34 | 34.76 | 0.00 | 17.60% | 16.98% |
| logMAR 0.51–1.0 | 45.75 | 39.02 | 0.00 | 42.29% | 32.83% |
| logMAR >1.00 | 37.91 | 21.95 | 100.00 | 19.53% | 10.57% |
| Diagnosis | | | | | |
| AMD | 70.20 | 41.36 | 44.49 | 0.00% | 34.72% |
| All other | 29.80 | 58.64 | 55.51 | 100.00% | 65.28% |
| Age | | | | | |
| Mean | 79.75 | 69.30 | 69.59 | 41.34 | 67.05 |
| SD | 8.12 | 15.71 | 17.85 | 16.69 | 19.38 |
| | Patient Characteristics of IVI 32 and 76 Data Sets | | | | |
| | Finger IVI 76 (n = 95) | Lamoureux-1 IVI 32 (n = 192) | Lamoureux-2 IVI 32 (n = 319) | | |
| Gender (%) | | | | | |
| Male | 35.79 | 35.42 | 36.68 | | |
| Female | 64.21 | 64.58 | 63.32 | | |
| VA[a] | | | | | |
| logMAR 0–0.3 | 0.00 | 11.46 | 14.15 | | |
| logMAR 0.31–0.5 | 0.00 | 26.04 | 27.99 | | |
| logMAR 0.51–1.0 | 34.74 | 51.56 | 46.54 | | |
| logMAR >1.00 | 65.26 | 10.94 | 11.32 | | |
| Diagnosis | | | | | |
| AMD | 40.22 | 61.46 | 53.77 | | |
| All other | 59.78 | 38.54 | 46.23 | | |
| Age | | | | | |
| Mean | 67.20 | 75.82 | 75.77 | | |
| SD | 19.18 | 13.03 | 12.93 | | |

[a] LogMAR 0–0.3 (Snellen >20/40); logMAR 0.31–0.5 (20/40–20/63); logMAR 51–1.0 (20/64–20/200); logMAR >1.00 (<20/200).

binocular viewing (some data sets precategorized VA; e.g., less than 20/400), ocular disorder diagnosis, date of data collection, date of birth or age, and gender.

The data from six principal investigators representing eight data sets of low vision patient responses to three different versions of the IVI (28-item, 32-item, and 76-item) were pooled for initial assessment. The IVI VFQ has undergone several iterations, but in all cases each item asks participants how much the visual impairment interfered with the performance of a specific activity in the past month. During the development and testing phase, there were 76 items, each with six response categories. The number of items was later reduced to 32 to make the IVI administration less burdensome to the patient.[20,21] Further analyses revealed problems with the item fit statistics and ordering of response category thresholds. To rectify these problems, four items were eliminated, reducing the IVI from 32-items to the current 28-item version, and number of response

categories was reduced from six to four with the exception of two items, for which response categories were reduced to three.

To equate the scales of the different questionnaires such that the higher person measure reflected greater ability, the rank scores of the IVI 32 and IVI 76 response categories were reversed from a disability to an ability scale to agree with the IVI 28 ability scale. Items with a different number of response categories were treated as separate item groups. To assess potential differences between populations in response to IVI items, and whether it would be reasonable to combine data sets, two initial differential item functioning (DIF) analyses were performed. The first DIF analyses were by study and center (1) 28-item questionnaire alone, and (2) combined 32-item and 76-item questionnaires (from the 76-item questionnaire, only the corresponding 32 items were included). Once it was determined that the DIF magnitude by study and center was acceptable, additional DIF analyses were performed for the 28-item data set to assess the impact of covariates (age, gender, VA, and diagnosis) on item estimates. Further refinement of the IVI 28 analyses occurred for items #14 and #15 depending on whether the investigator used three or four response categories.

## Assessing Internal and External Validity

Rasch analysis using the Andrich Rating Scale Model with item grouping (Winsteps version 3.65) was performed on the 28-item data set alone and the combined 32- and 76-item data sets to estimate person and item separation reliability, item measures and response category thresholds, and a person measure for each patient. To examine whether and how much centers varied from each other in their response patterns to the difficulty of each of the items, DIF contrast was calculated in the 28-item data set alone and the combined 32-item and 76-item data set, using the Mantel Haenszel Test with Bonferroni correction. Ninety-five percent confidence intervals on DIF were applied to further evaluate the possibility of DIF by data set.

All subsequent analyses were performed with the IVI 28-item, as the number and description of response categories between the 28-item IVI differ from the combined 32-item and 76-item IVI, and the 28-item is the version in active use. The person and item measure distributions from responses to the IVI 28-item were mapped to assess precision of the IVI. Accuracy of the measure was assessed using information-weighted mean square fit statistics (infit mean square) transformed to $z$ scores.[22] These transformed infit mean-squares have an expected value of 0 and an expected variance of 1.0. The frequency of infit mean squares for persons was binned into 0.05 increments and compared to the corresponding $\chi^2$ divided by degrees of freedom (number of items minus 1) probability mass function.

Item estimates were then anchored to the initial IVI-28 item analysis to perform a second set of Rasch analyses using item subsets representing each of the IVI subscales (i.e., reading, mobility, and emotional items), which originally were defined during the development and refining of the IVI.[18] Principal axis factoring on the four sets of person measures representing the three subscales and the overall measure was performed to assess dimensionality of the IVI. To evaluate external validity, DIF analysis and analysis of variance (ANOVA) were used to examine the effects of covariates on item estimates, including age, VA, gender, and diagnosis. Here, the objective of the DIF analysis was to examine if estimated measures for different items varied between subgroups of patients (categorized by VA, age, etc.). Age was binned into four categories (18–40 years, 41–60 years, 61–80 years, >80 years); VA was binned into four categories (logMAR 0.0–0.3, logMAR 0.31–0.5, logMAR 0.51–1.00, logMAR >1.0), and diagnosis was binned into two categories, AMD and other.

## Results

Patient characteristics, visual impairment measures, and responses to the IVI were obtained from eight study samples totaling 2867 patients tested between years 2000 and 2014 (Table 1). The age range was 18 to 103 years (mean = 59 years; SD = 22; median = 62 years). AMD was the diagnosis in 29% of the sample, while the remaining 71% represented all other ocular disorders diagnoses. Better eye or binocular VA ranged from Snellen 20/20 (logMAR 0.00) to no light perception. Data sets from Finger et al.[13] for the IVI-28 and IVI-76 were obtained mostly from patients with severe to profound VA loss (100% to 65%, respectively), while the Gothwal et al., data set was composed mostly of younger patients (mean age 41 years), males (76%), with no cases of AMD.[17]

Item measure estimates for the IVI 28-item (Fig. 1a) and combined IVI 32-item and 76-item (Fig. 1b) revealed good overall reliability ($R^2 = 0.99$ for both), with the majority of items showing data set-related DIF magnitude less than 0.5 logit relative to the average item measure for combined data sets. When
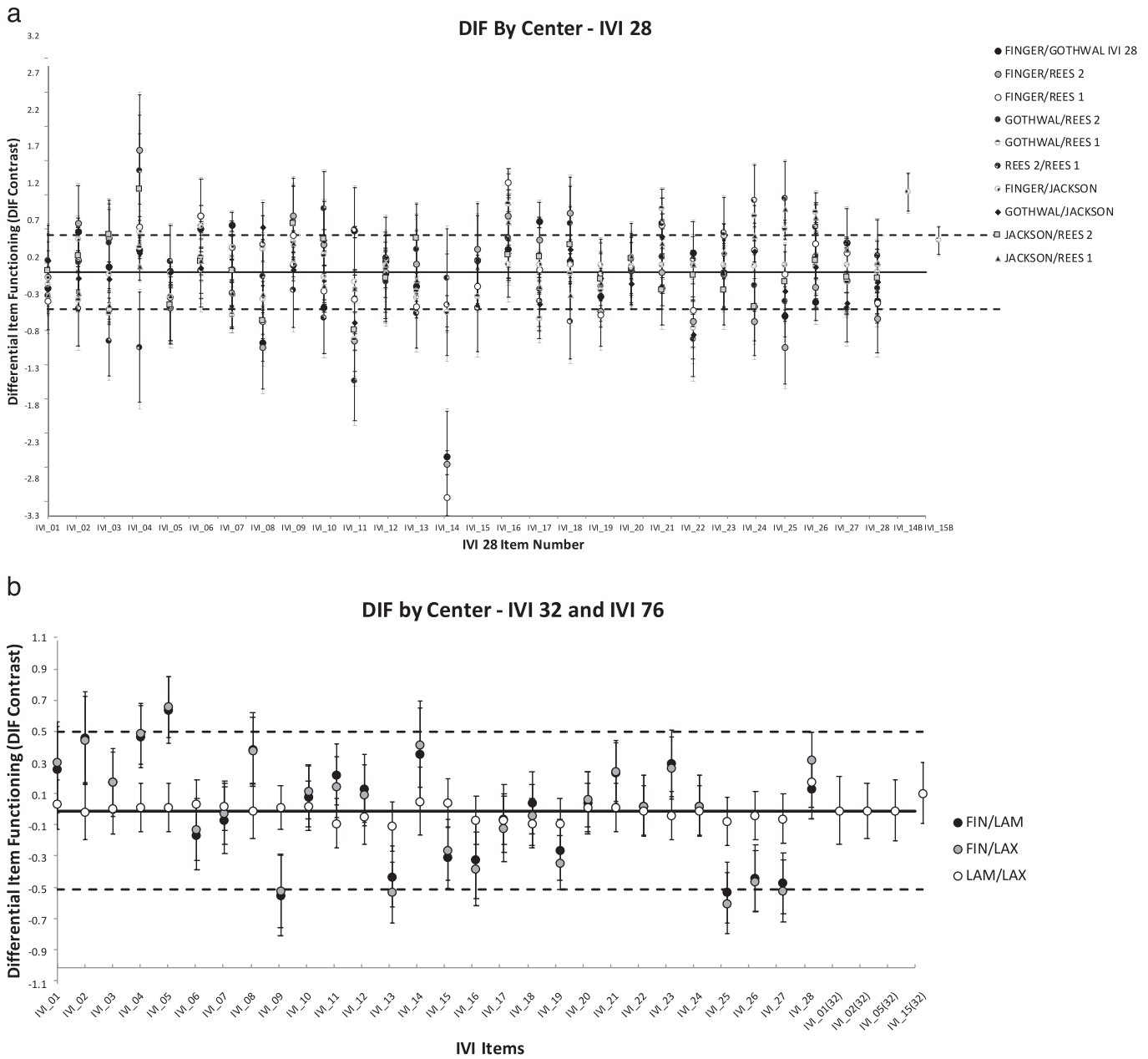
a

**DIF By Center - IVI 28**

b

**DIF by Center - IVI 32 and IVI 76**

**Figure 1.** (a) Illustrates a Bland-Altman type plot of item measures comparing item estimates for the 28-item data set. The 30 items on the *x*-axis represent the 28 items from the IVI 28 for which each item has four response categories, and a replication of two items, #14 and #15 (listed as #14B and 15B), which have three response categories (these two items were modified during IVI development). (b) Compares item estimates of each of the IVI-32 and 76-item versions of the instrument. Only the 32 corresponding items from the IVI 76 were included in the analysis. Zero corresponds to the item measure estimates from all of the data sets combined, with each center's difference from the combined item measures plotted with 95% confidence intervals.

combining the two 32-item data sets and 76-item data set of the IVI, only item #5 and #25 ("eyesight interfering with recognizing and meeting people" and "worried about your eyesight getting worse," respectively) exceeded DIF magnitude > +0.5 logits. A greater number of items among comparisons of the five IVI-28 data sets yielded DIF > +/− 0.5 logits,

including reading-related items (#8-reading labels or instructions on medicine, #14 and #14B-reading ordinary size print), and emotional items (#22-frustrated or annoyed because of eyesight, #24-felt sad or low because of eyesight, #25-worried about your eyesight getting worse). Items #14B and #15B (Fig. 1a) are unique as the number of response
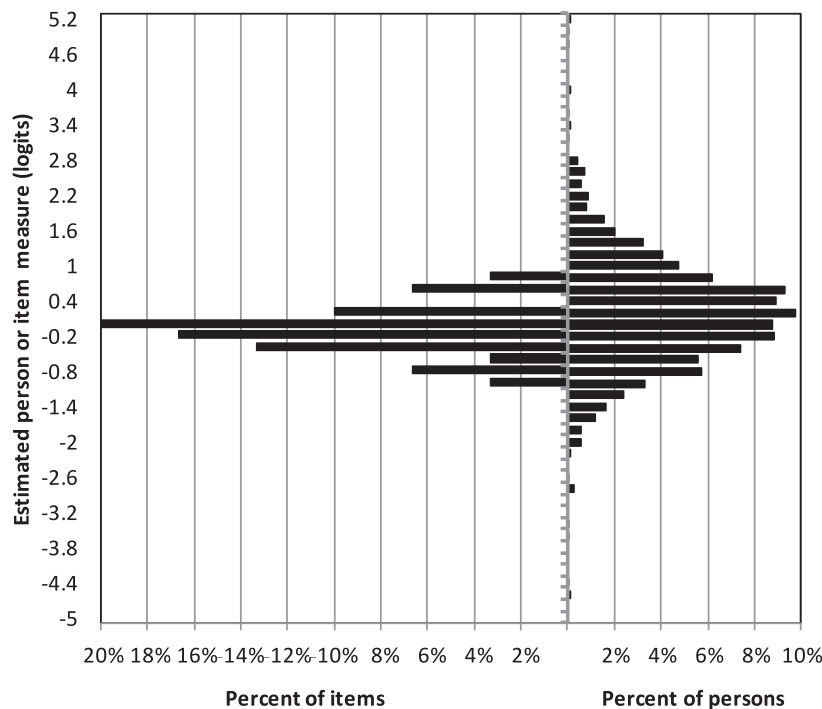
**Figure 2.** Histograms of person measures (*black bars* to right of vertical axis) and item measure (*black bars* to left of vertical axis) estimates from responses to all items of the IVI-28.

categories for these two items was modified to three during the development of the IVI to eliminate the occurrence of disordered thresholds with the Andrich rating scale model. This is in contrast to all other items that contained four response categories.

Comparison of person and item measure distributions on a logit visual ability scale is illustrated in Figure 2. IVI-28 item measures range from −0.87 to 0.95 logits and person measures range from −4.8 to 5.12 logits. The item measure distribution is fairly well centered on the person measure distribution (mean = 0.0 logit, SD = 0.4 logit); however, near center of the distribution there is some loss of precision evident at 0.4 logit. At the extremes of the scale, there are no items targeted to patients with milder or more severe loss in visual ability.

The person and item infit mean-squares are expected to be distributed as $\chi^2$/degrees of freedom if the observed pattern of ratings across persons and items agree with Rasch measurement model assumptions—a test of internal validity of the estimated person and item measures, respectively.[12] A histogram of person infit mean squares is illustrated in Figure 3 and compared to the expected $\chi^2$ normalized to its degrees of freedom probability mass function (red curve). There is good agreement between the

observed distribution of person measure infit mean squares (histogram) and model expectations (probability mass function represented by the red curve), but the observed distribution being broader (less kurtosis) than the expected distribution.

The IVI-28 item measure estimates, standard errors of the estimate and fit statistics, both as infit mean square residuals and as standard normal deviates are displayed in Table 2. "Reading ordinary size print (for example newspapers)" is the item requiring the greatest visual ability (0.95 and 0.72 logits for 14 and 14B, respectively), and "general safety at home" requires the least visual ability (−0.87 logit). Findings from the z-score distribution of the IVI-28 infit mean square residuals show large departures from the expected distribution, evidenced by z-scores ranging from −9.9 to +9.9 modeled SDs around the expected value of zero (empirical mean of z-scores = −0.3, SD = 5.7), which most likely indicates that there are multiple sources of variance in the observations and that the estimated measures are not unidimensional. (n.b., If there were a single random source of variance, the expected distribution would have a mean of 0 and a SD of 1.)

To explore dimensionality of the IVI-28 at the item level and better understand sources of variance in
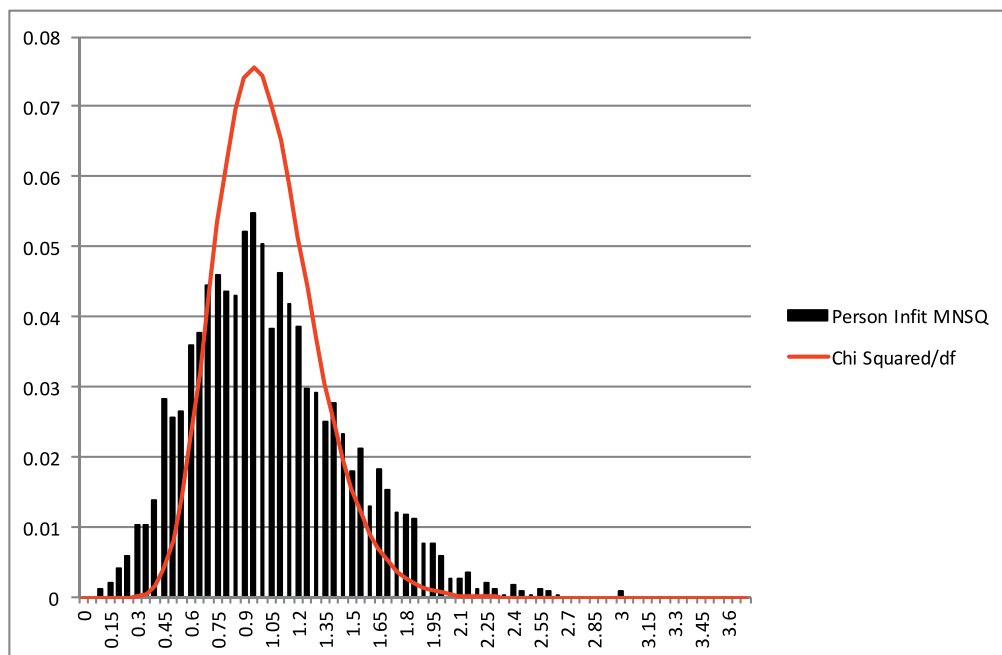
**Figure 3.** Histograms of person infit mean squares measures (*black bars*) compared to the expected $\chi^2$ normalized to its degrees of freedom probability mass function (*red curve*).

person measure estimates, principal axis factoring was employed on person measure estimates from responses to four subsets of items in the IVI-28: the three predefined domains (reading, mobility, emotional) and all 28 items. The person measures for the three domains correlated highly with the principal axis (reading $r = 0.79$, emotion $r = 0.79$, mobility $r = 0.90$), which is defined by the person measures estimated from responses to all 28 items.

DIF analysis by gender revealed 7 out of the 28 items with statistically significant DIF ($P < 0.002$; alpha of 5% divided among the 28 items); however, DIF magnitude for all comparisons was $< 0.5$ logit. DIF by disorder (AMD versus other) was significant for 15 out of the 28 items ($P < 0.002$); however, only 3 out of the 28 items had DIF $> 0.5$ logit. DIF by age was significant ($P < 0.0003$) for 50 (28%) of the 168 comparisons (28 items $\times$ 6 age comparisons). DIF by VA was significant ($P < 0.0003$) for 40 (22%) of the 168 comparisons; however, DIF for VA comparisons only exceeded 0.5 logits in 11% of comparisons. Across all IVI-28 DIF analyses, the greatest DIF magnitude observed (1.23 logits) was for item #14 (reading ordinary size print) for comparison between VA category 3 (logMAR 0.51–1.00) and 4 (logMAR $>1.0$).

ANOVA using the four covariates showed VA ($P$ $< 0.001$) and gender ($P < 0.002$) were predictors of visual ability. Person measure declined monotonically with each worsening VA category. Females showed a relatively worse person measure across items. Ocular disorder diagnosis was not statistically significant ($P = 0.173$), and age had no meaningful effect ($P = 0.741$) on person measure estimates.

## Discussion

Findings from validity testing of the IVI show that we can combine disparate data sets of low patient responses from multiple studies to calibrate item measures and response category thresholds (Table 2 and Supplementary Material) for the research and clinical community to use when employing the IVI as a PROM. Consistent with similar efforts in other fields standardizing the unit of measurement ensures invariant scales across studies that employ the same instrument, enabling users to administer the IVI to any size sample of visually impaired patients and directly compare results across studies.[9]

Previous studies have divided the IVI items into three domains: reading, mobility, and emotional health. However, our analysis demonstrates that most of the variance in the observed person measures (75%) can be explained by a single vision-related principal

axis that is highly correlated with reading and mobility. Given the high correlations observed and the majority of variance explained by vision-related items, a summary IVI score representing all items (in contrast to item subset analyses; e.g., domains), is likely adequate when measuring and reporting outcomes from medical interventions. Medical interventions are typically restorative in nature and the treatment effect is at the person level. This is in contrast to low vision rehabilitation where the effects are observed at the item level.[23] People's observations to the IVI reasonably fit the expectations of the model, however, like other VFQs, item fit analysis showed large amounts of variance.[24,25] DIF by center showed mostly good agreement between data sets in item measure estimates. Data sets (Finger et al.[13] and Gothwal et al.[17]) with extreme demographic (younger) and vision characteristics (more profound levels of VA loss) were the greatest sources of DIF between centers. ANOVA by covariate (age, gender, disorder, VA) showed that certain items more commonly showed greater DIF magnitude. Item examples included eyesight interfering with reading ordinary size print (for example, newspapers), reading labels or instructions on medicine bottles, go carefully to avoid falling or tripping, felt sad or low because of your eyesight, worried about your eyesight getting worse, visiting family or friends, and opening packages. Some of the bias evident in responses to these items may represent true underlying differences between data sets, such as an excess of extremes in category responses. This observation does not preclude use of the IVI. Rather, were we designing a new instrument, we would consider modifying these items to minimize DIF for the relevant covariates differing between data sets.

The IVI person-item map shows that the strongest information in the estimated measures occurs in the middle of the visual ability distribution. At both tails of the distribution, there is an absence of items targeting patients with either very good or very poor visual ability. This is an important consideration when determining how appropriate the IVI questionnaire will be for measuring baseline visual ability and treatment outcomes in a given population. Detecting meaningful change (e.g., minimum clinically important difference) with a VFQ requires consideration of larger standard errors with extreme scores, and therefore a smaller, but meaningful effect may not be resolvable in people close to the floor or ceiling of the measure. To better target visual ability measurement in populations with more profound vision loss,

**Table 2.** IVI-28 Item Measures, Standard Errors, Infit Mean Squares, and Z-Score Distributions

| IVI Item | Item Measure | Standard Error | Infit Mean Square | Infit z-Score |
|---|---|---|---|---|
| IVI 01 | 0.19 | 0.03 | 1.05 | 1.74 |
| IVI 02 | 0.02 | 0.03 | 0.99 | −0.45 |
| IVI 03 | 0.2 | 0.03 | 0.86 | −5.31 |
| IVI 04 | −0.17 | 0.02 | 0.98 | −0.62 |
| IVI 05 | 0.18 | 0.02 | 1.66 | 9.9 |
| IVI 06 | −0.35 | 0.02 | 1 | 0.16 |
| IVI 07 | −0.62 | 0.03 | 1.11 | 3.63 |
| IVI 08 | 0.8 | 0.03 | 1.48 | 9.9 |
| IVI 09 | −0.26 | 0.02 | 1.04 | 1.39 |
| IVI 10 | 0.09 | 0.02 | 0.86 | −5.19 |
| IVI 11 | 0.1 | 0.02 | 1.03 | 1.05 |
| IVI 12 | 0.05 | 0.02 | 0.89 | −4.21 |
| IVI 13 | 0.01 | 0.02 | 0.9 | −4.07 |
| IVI 14 (3 resp cat) | 0.95 | 0.04 | 1.36 | 9.82 |
| IVI 14B (4 resp cat) | 0.72 | 0.06 | 1.07 | 1.06 |
| IVI 15 (3 resp cat) | −0.01 | 0.04 | 0.89 | −3.98 |
| IVI 15B (4 resp cat) | 0.04 | 0.05 | 0.91 | −1.77 |
| IVI 16 | −0.87 | 0.03 | 1.34 | 9.59 |
| IVI 17 | −0.78 | 0.03 | 1.04 | 1.33 |
| IVI 18 | −0.16 | 0.02 | 0.8 | −7.82 |
| IVI19 | 0.15 | 0.02 | 0.76 | −9.9 |
| IVI 20 | 0.13 | 0.02 | 0.71 | −9.9 |
| IVI 21 | −0.24 | 0.02 | 1.16 | 5.58 |
| IVI 22 | 0.22 | 0.02 | 0.93 | −2.88 |
| IVI 23 | −0.34 | 0.02 | 1.03 | 1.19 |
| IVI 24 | −0.06 | 0.02 | 0.95 | −1.81 |
| IVI 25 | 0.22 | 0.02 | 1.05 | 1.89 |
| IVI 26 | 0 | 0.02 | 0.89 | −4.15 |
| IVI 27 | −0.44 | 0.02 | 1.14 | 4.9 |
| IVI28 | 0.22 | 0.02 | 0.76 | −9.9 |
| IVI-28 Response category thresholds (4 response categories)[a] | | | | |
| 1 | −0.12 | | | |
| 2 | −0.43 | | | |
| 3 | 0.55 | | | |
| IVI-28 Response category thresholds for item 14B and 15B with 3 response categories[b] | | | | |
| 1 | −0.4 | | | |
| 2 | 0.4 | | | |

Resp cat, response categories.
[a] Average functional reserve in equated measurement units for each response category (0 = a lot to 3 = not at all).
[b] During modification of IV-28 items 14 and 15 were collapsed into three categories.

the IVI –Very Low Vision and the Ultra-Low Vision Visual Functioning Questionnaire were developed.[13,26]

VA was the strongest predictor of IVI person measure estimates with evidence of a monotonic decline in person measures with declining VA. This is consistent with all valid VFQs and supports the premise that VA (or resolution capacity) is one of the primary factors in assessing visual ability.[27] Gender was also predictive of IVI person measure estimates, with females showing relatively worse person measure estimates compared to males. Neither ocular diagnosis (AMD versus other) nor age were associated with person measure estimates.

The IVI-28 item calibration estimates provided in Table 2 or in Supplementary File S1 (MS excel with computations encoded or Supplementary File S2 Winsteps CON file) may be used by researchers and clinicians to measure visual ability in patients with vision impairment. The strengths of this work include the large number of respondents to the IVI and the diverse patient population, both geographically and visually. Providing these IVI item calibrations enables researchers to more accurately and precisely compare treatment outcomes involving people with vision impairment, especially when the number of patients is small. Study limitations include an inability to assess responses to IVI items by more detailed diagnoses and that the calibrations do not reflect responses from individuals younger than 18 years old.

## Conclusion

Researchers and clinicians using the IVI questionnaire to assess vision-related patient reported outcomes can use the supplied 28-item calibrated item and threshold measures to better enable direct comparisons across clinical trials and on an individual patient basis. Internal and external validity of the IVI questionnaire shows that it best targets people with vision impairment in the middle of the visual ability spectrum, with baseline person measure estimates most strongly predicted by VA and less, but significantly so by gender.

## Acknowledgments

## References

1. Massof RW, Rubin GS. Visual function assessment questionnaires. *Surv Ophthalmol*. 2001;45:531–548.
2. Margolis MK, Coyne K, Kennedy-Martin T, Baker T, Schein O, Revicki DA. Vision-specific instruments for the assessment of health-related quality of life and visual functioning: a literature review. *Pharmacoeconomics*. 2002;20:791–812.
3. de Boer MR, Moll AC, de Vet HC, Terwee CB, Volker-Dieben HJ, van Rens GH. Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic Physiol Opt*. 2004;24:257–273.
4. Khadka J, McAlinden C, Pesudovs K. Quality assessment of ophthalmic questionnaires: review and recommendations. *Optom Vis Sci*. 2013;90:720–744.
5. Centers for Medicare and Medicaid Services. Functional reporting: PT, OT, and SLP services. Available at: https://www.cms.gov/Medicare/Billing/TherapyServices/Functional-Reporting.html. Accessed May 1, 2018.
6. Methodology Committee of the Patient-Centered Outcomes Research Institute (PCORI). Methodological standards and patient-centeredness in comparative effectiveness research: the PCORI perspective. *JAMA*. 2012;307:1636–1640.
7. Food and Drug Administration. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. Available at: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. Accessed May 1, 2018.
8. Chan TL, Perlmutter MS, Andrews M, et al. Equating visual function scales to facilitate reporting of medicare functional G-code severity/complexity modifiers for low-vision patients. *Arch Phys Med Rehabil*. 2015;96:1859–1865.
9. Garcia SF, Cella D, Clauser SB, et al. Standardizing patient-reported outcomes assessment in cancer clinical trials: a patient-reported outcomes measurement information system initiative. *J Clin Oncol*. 2007;25:5106–5112.
10. Rosenfeld PJ, Brown DM, Heier JS, et al. Ranibizumab for neovascular age-related macu-

lar degeneration. *N Engl J Med*. 2006;355:1419–1431.

11. Hudson HL, Lane SS, Heier JS, et al. Implantable miniature telescope for the treatment of visual acuity loss resulting from end-stage age-related macular degeneration: 1-year results. *Ophthalmology*. 2006;113:1987–2001.

12. Massof RW. Understanding Rasch and item response theory models: applications to the estimation and validation of interval latent trait measures from responses to rating scale questionnaires. *Ophthalmic Epidemiol*. 2011;18:1–19.

13. Finger RP, McSweeney SC, Deverell L, et al. Developing an instrumental activities of daily living tool as part of the low vision assessment of daily activities protocol. *Invest Ophthalmol Vis Sci*. 2014;55:8458–8466.

14. Ang M, Man R, Fenwick E, Lamoureux E, Wilkins M. Impact of type I Boston keratoprosthesis implantation on vision-related quality of life. *Br J Ophthalmol*. 2017;102:878–881.

15. Casparis H, Lindsley K, Kuo IC, Sikder S, Bressler NM. Surgery for cataracts in people with age-related macular degeneration. *Cochrane Database Syst Rev*. 2017;2:CD006757.

16. Fenwick EK, Cheung GCM, Gan AT, et al. Change in vision-related quality of life and influencing factors in Asians receiving treatment for neovascular age-related macular degeneration. *Br J Ophthalmol*. 2017;102:948–953.

17. Gothwal VK, Bharani S. Outcomes of multidisciplinary low vision rehabilitation in adults. *Invest Ophthalmol Vis Sci*. 2015;56:7451–7461.

18. Lamoureux EL, Pallant JF, Pesudovs K, Rees G, Hassell JB, Keeffe JE. The effectiveness of low-vision rehabilitation on participation in daily living and quality of life. *Invest Ophthalmol Vis Sci*. 2007;48:1476–1482.

19. Gothwal VK, Bagga DK. Vision and quality of life index: validation of the Indian version using Rasch analysis. *Invest Ophthalmol Vis Sci*. 2013;54:4871–4881.

20. Weih LM, Hassell JB, Keeffe J. Assessment of the impact of vision impairment. *Invest Ophthalmol Vis Sci*. 2002;43:927–935.

21. Lamoureux EL, Pallant JF, Pesudovs K, Hassell JB, Keeffe JE. The impact of vision impairment questionnaire: an evaluation of its measurement properties using Rasch analysis. *Invest Ophthalmol Vis Sci*. 2006;47:4732–4741.

22. Wilson EB, Hilferty MM. The distribution of chi-square. *Proc Natl Acad Sci U S A*. 1931;17:684–688.

23. Massof RW, Stelmack JA. Interpretation of low-vision rehabilitation outcome measures. *Optom Vis Sci*. 2013;90:788–798.

24. Massof RW, Ahmadian L, Grover LL, et al. The activity inventory: an adaptive visual function questionnaire. *Optom Vis Sci*. 2007;84:763–774.

25. Massof RW. An interval-scaled scoring algorithm for visual function questionnaires. *Optom Vis Sci*. 2007;84:689–704.

26. Jeter PE, Rozanski C, Massof R, Adeyemo O, Dagnelie G; PLOVR Study Group. Development of the ultra-low vision visual functioning questionnaire (ULV-VFQ). *Transl Vis Sci Technol*. 2017;6:11.

27. Goldstein JE, Chun MW, Fletcher DC, Deremeik JT, Massof RW; Low Vision Research Network Study Group. Visual ability of patients seeking outpatient low vision services in the United States. *JAMA Ophthalmol*. 2014;132:1169–1177.

Author/s:

Goldstein, JE; Fenwick, E; Finger, RP; Gothwal, V; Jackson, ML; Lamoureux, E; Rees, G; Massof, R

Title:

Calibrating the Impact of Vision Impairment (IVI): Creation of a Sample-Independent Visual Function Measure for Patient-Centered Outcomes Research

Date:

2018-12-01

Citation:

Persistent Link:

http://hdl.handle.net/11343/253676

File Description:

Published version

License: